WHAT DETAILS DO GEOMETRY TEACHERS EXPECT IN STUDENTS' PROOFS? A METHOD FOR EXPERIMENTALLY TESTING POSSIBLE CLASSROOM NORMS

Justin K. Dimmel, Patricio G. Herbst

University of Michigan

We report on the development and piloting of a method that provides a sufficient condition for confirming that an observed regularity in a classroom is a norm. The method we describe is a refinement of the breaching experiment technique (Garfinkel, 1963; Mehan, 1979) that uses random assignment to experimental conditions as a means to facilitate controlled comparisons between participants' reactions to different episodes of instruction. We use this method to confirm the existence of normative ways that teachers scrutinize the details of proofs in geometry.

INTRODUCTION

International comparisons of teaching have brought attention to the notion of cultural scripts and the claim that regularities are observed across episodes of teaching in a given country (Stigler & Hiebert, 2009; Santagata & Stigler, 2000). The existence of these cultural scripts is warranted by observations of different teachers who share a national culture engaging in stable patterns of classroom activity—patterns that are similar to each other yet distinct from patterns of teachers from other national cultures (Stigler & Hiebert, 2009). On account of the scale of such comparisons, the identified scripts have been largely subject-independent and rather general. Furthermore, the extent to which cultural scripts capture *norms* of classroom action—that is, what is expected to happen, for the absence of which would be seen as a violation of the social order (Garfinkel, 1963)—as opposed to provide descriptions of what is observed to happen in classrooms—is an open question. A social norm is not merely an action that might be frequently observed, but actually an action that participants expect (or expect their coparticipants) to engage in. Developing methods for identifying the classroom regularities that are actually norms is pressing because providing an account of what teachers expect to happen in classrooms—as opposed to just recording those things that do happen—brings us closer to understanding what it might cost to change classroom instruction.

In this paper, we report on the development and piloting of a method that provides a sufficient condition for confirming that an observed regularity in a classroom is a norm. We use for that the classroom activity *doing proofs in geometry* (Herbst & Miyakawa, 2008) and norms that we call *semiotic norms*. By semiotic norm, we mean a norm of the way in which semiotic resources (e.g., written words, diagrams) are used to produce and evaluate mathematical work. The method we describe is a refinement of the breaching experiment technique (Garfinkel, 1963; Mehan, 1979) that uses

2014. In Nicol, C., Liljedahl, P., Oesterle, S., & Allan, D. (Eds.) Proceedings of the Joint Meeting of PME 38 and PME-NA 36,Vol. 2, pp. 393-400. Vancouver, Canada: PME.

random assignment to experimental conditions as a means to facilitate controlled comparisons between participants' reactions to different episodes of instruction. The method we developed for confirming the existence of classroom norms will help researchers describe more precisely the mathematics that students have an opportunity to learn and will also help identify levers for piecemeal alterations to curriculum and instruction in order to improve the mathematical quality of the work students are involved in.

THEORETICAL FRAMEWORK

Doing proofs in geometry is an example of an *instructional situation*: A stable segment of classroom activity within which students trade (or "cash") completed work for a claim—from the teacher—that they have acquired a particular item of knowledge (Herbst 2006; Herbst & Chazan, 2011). When doing proofs, the work to be produced is a proof of a particular mathematical statement and when a proof is so produced it may be exchanged (i.e., cashed) for a claim that some knowledge exists implicit in that proving work (such as the knowledge of how to produce a specific kind of mathematical argument). Within any instructional situation, the exchange of work for knowledge-claims is made possible through the available semiotic resources (Herbst & Chazan, 2012) that, together, comprise the semiotic currency of the situation. We are concerned with describing the normative ways that semiotic resources are used in such situations, or what we call semiotic norms.

From the perspective of social semiotics (van Leeuwen, 2004), instructional situations may be conceptualized as *genres* of classroom activity that have different realizations (Christie, 1997; Lemke, 1990; Martin & Rose, 2008). From video records of different geometry classrooms doing proofs[1], we identified presenting/checking a proof as a realization of the doing proofs situation in which the teacher presents a completed proof to the class and the students in the class take turns scrutinizing its written steps. In video episodes of different geometry lessons, there were recurring instances of details of the proof being insufficient under such scrutiny. These included instances when conceptual entailments—such as the conclusion that two angles that form a linear pair are supplementary—were not unpacked into more primitive steps (i.e., a statement that identifies such angles as being a linear pair followed by a statement that angles forming a linear pair are supplementary) and instances when distinctions between geometric objects and their measures (such as a segment versus the length of a segment) were not strictly enforced. Since the kinds of details that were scrutinized in the written arguments of proofs recurred in different geometry classrooms, we hypothesized the existence of a *details norm* when checking proofs. To confirm that there are, in fact, normative ways of scrutinizing the details of a proof, we devised a planned comparison study between groups of teachers in treatment and control conditions. The design of the experiment and the results of the data analysis are reported in the next sections.

[1] This data corpus had been gathered with the support of NSF grant 0133619 to P. Herbst.

METHOD

The method we developed to confirm the existence of the details norm combines the technique of a virtual breaching experiment (Herbst & Chazan, 2003; Nachlieli & Herbst, 2009) with a planned comparison study. As a virtual breaching experiment, we developed storyboards consisting of a sequence of classroom images to represent episodes of high school geometry lessons and showed these to participants. There were from 9 to 22 images in each storyboard. These scripted image sequences were adaptations of geometry lessons that were based on video recordings of classrooms doing proofs. As a planned comparison study, participants were randomly assigned to treatment and control conditions in which the teacher in the storyboard episode breaches (treatment) or complies with (control) the details norm. The purpose of randomly assigning participants to conditions was to be able to compare reactions (both within and across conditions) to the different lessons.

We used image sequences, rather than actual video, for two reasons. One, since we wanted to compare reactions to episodes where a norm is breached to reactions to episodes where a norm is not breached, using actual classroom video was not feasible, since in actual geometry classrooms, the norm is not usually breached. Second, using sequences of images allowed the breach and control conditions to feature representations of instruction that were minimally different from each other—that is, for a given instructional episode, its breach and control versions were identical except for those images in the sequence that depicted the breach of (or compliance with) the norm. This principle of minimal variation allowed us to make comparisons across the conditions.

What is described above as the details norm was the subject of four classroom stories (A, B, C, D), and each of these classroom stories had a version (A', B', C', D') in which the norm was breached and a version in which the norm was not breached. In stories A and B, the teacher allows minor omissions[2] in the written argument of a proof to stand without correction (thus breaching the norm), while in stories A' and B'—the control duals of A, B, respectively—the teacher corrects the omissions. In stories C and D, the teacher insists that students explicitly justify claims[3] that are tacitly warranted by a diagram (thus breaching the norm), while in stories C' and D'—the control duals of C and D, respectively—the teacher uses the diagram to elide some steps in the proof.

As a group, these four sets of stories concern the necessary details of the semiotic currency for a valid exchange of proof-for-credit when doing proofs. We hypothesize that the teacher in stories A and B would be seen as breaching the details norm because the teacher accepts less detail in the written argument of a proof than what is usually

[2] Respectively: failing to include an explicit step that establishes the congruence of two segments from the definition of midpoint, and failing to distinguish between angles and their measures.

[3] Respectively: that a point of intersection between two rays exists, and that two angles are collinear.

required, while the teacher in stories C and D would be seen as breaching the details norm because the teacher asks for more details than what is usually required. The instrument we developed thus allowed us to test two different ways in which the details provided in a proof might be seen as breaching the norm. We thought important to test both hypotheses to be able to argue that the norm is not actually a generic one (insisting on detail, no matter what detail), but rather a mathematically specific one—some details are insisted upon, others frowned upon, and the semiotic systems involved are the bearers of the distinction.

The structure of the instrument was the same for all stories: participants were shown one of the classroom stories, then asked a series of questions. These included a general open response question—"What did you see happening in this scenario?"—a general closed-response rating question—"How appropriate was the teacher's review of the proof?"—and two targeted, closed-response rating questions (described below). All of the rating questions used the same 6-point Likert-style rating scale, with choices from 1 (very inappropriate) to 6 (very appropriate). The rating questions also included a "please explain your rating" follow-up prompt.

For the targeted rating questions, participants were shown a "clip" of the story (that is, a segment of the storyboard) that focused on a particular teaching action. One of these targeted rating questions showed participants the 3 to 5 image clip in which the norm was either breached or not breached, stratified by condition. The purpose of this targeted rating question was to focus participants' ratings on the part of the story where the teacher complies with or departs from the norm. Participants were also asked to rate a different clip. For this other targeted rating question, participants in the breach/control conditions were shown identical sets of 3-5 images in which the teacher in the story does a routine instructional action unrelated to the target norm. It was possible to identify such clips because each set of breach/control stories were identical except during those parts of the story that represent the breach of (or compliance with) the target norm. The purpose of including the two types of targeted rating questions was to enable comparisons across the breach and control conditions. These comparisons and their results are described below.

DATA

We gathered data from 34 high school teachers (working in different schools and districts) during a pilot study in the fall of 2013. The teachers were randomly assigned to treatment and control conditions. 16 teachers were assigned to a condition where they viewed stories that breached the norm (7 assigned to the "less details" breach, 9 assigned to the "more details" breach), and 18 teachers were assigned to a condition where they viewed stories that complied with the norm (9 assigned to the "less details" control, 9 assigned to the "more details" control). Within each condition, a teacher either viewed two stories that breached the norm or two stories that complied with the norm. No participant viewed the breach and control version of the same story, and the order in which the stories appeared was randomized (to neutralize any effects from the

order in which the stories are viewed). Since each participant viewed and rated two stories, there were 32 responses to each question about stories where the target norm was breached and 36 responses to each question about stories where a norm was not breached.

ANALYSIS AND RESULTS

The study was a planned comparison study between participants assigned to treatment and control conditions. Since a norm is not only what is routine but also what is expected, we hypothesized that participants would find the work of the teacher less appropriate in those stories that breached the hypothetical norm. We made three comparisons of answers on closed-response questions both across and also within the different conditions to test this hypothesis. The first was a comparison of the mean scores on the general rating question across the breach and control conditions. The second was a comparison of ratings on the targeted rating questions between breach conditions and control conditions, while the third was a within-condition comparison between ratings on the targeted rating questions—i.e., comparing ratings on the breach/nonbreach storyboard segments to the ratings on the other storyboard segment within a condition. These comparisons and the results of statistical tests are reported below.

Comparison 1: Across condition comparison of mean scores on the general rating questions

The general closed-ended rating question asked participants to rate the appropriateness of the teacher's review of the proof: "how appropriate was the teacher's review of the proof?" There were 32 responses to this question across 4 stories that breached a norm, and 36 responses to this question across 4 stories that complied with a norm. Using the 6-point Likert-syle rating scale for appropriateness described above, the mean rating of the breach responses was 1.14 points lower than the mean rating of the control response (3.47 compared to 4.61, respectively), a statistically significant difference in means at the .05 level (two-tailed, heteroscedastic t-test assuming unequal *Ns*, $p<.01$). Because of the random assignment, this difference in means provides some evidence that any secondary math teacher would notice when the details norm is breached when doing proofs in geometry.

Comparison 2: Across condition comparison of mean scores on the targeted rating questions

The targeted rating questions asked participants to rate the appropriateness of the teacher's actions at a specific place in the story. Each participant answered two types of targeted rating questions: one that targeted the place in the story where the teacher breaches (or complies with) the norm, and one that targeted a moment in the story when the teacher engages in some other action. We refer to the first type of targeted rating question as the "targeted breach/compliance" (TBC) rating and the second as the "targeted distracter" (TD) rating. By design, the TD rating questions targeted an action

that appeared in both the breach and compliance versions of a story, so participants across the conditions viewed identical story segments when answering this rating question. The purpose of including these targeted rating questions was to be able to compare ratings both across and within conditions at specific points in the stories.

Across the conditions (32 and 36 respective responses, as above), the mean ratings on the TBC questions for those who viewed breach stories was 1.61 points lower than the mean rating on the TBC questions for those who viewed compliant stories (2.78 to 4.39, respectively), a statistically significant difference at the .05 level (two-tailed, heteroscedastic *t*-test assuming unequal *Ns*, $p < .001$). This significant difference in means on the rating questions that target the moments in the stories that either breach or comply with the norm is complemented by a non-significant difference in means on the TD rating questions: 4.1 (breach) to 4.6 (control), a .5 difference that is not significant at the .05 level (two-tailed, heteroscedastic *t*-test assuming unequal *Ns*, $p = .13$). The significant difference in mean TBC ratings together with the non-significant difference in TD ratings suggests that participants' overall lower ratings on the breach stories (compared to the control stories, reported above) are linked to the teacher's breach of the norm, rather than some other action the teacher takes in the story. The experimental design and the deliberate scripting of the stories to be identical in all places except for where the teacher breaches the norm underscores this point.

Comparison 3: Within condition comparison of mean scores on the targeted rating questions

Further evidence that participants were responding to breaches of a norm—as opposed to other aspects of the stories—comes from within condition comparisons of the targeted rating questions (32 and 36 responses, as before). For the breach stories, the mean scores on the TBC ratings was 1.31 points lower than the mean scores on the TD ratings (2.79 to 4.1, respectively), a statistically significant difference at the .05 level (paired, two-sample *t*-test, $p < .001$). Complementing this, there was no significant difference between TBC and TD ratings for the stories in the control condition (means scores of 4.39 and 4.6, respectively, $p = .15$). The fact that, in the breach condition, participants' ratings on the TBC questions were significantly lower than their ratings on the TD questions—together with the fact that there were no such significant differences between the targeted rating questions for participants in the control condition—indicates that participants noticed the moments in the episodes of instruction when teachers were shown departing from the norm.

Open-response data

The open response data also indicate what participants view as appropriate or inappropriate ways of scrutinizing a proof. For example, a participant who viewed story D—one in which the teacher breaches the details norm by problematizing the existence of a point of intersection for the angle bisectors of a parallelogram—remarked: "The rays [of the parallelogram] intersect by definition. We don't need a theorem to justify it (participant ID 2248)". As a comment on this same

story, another participant remarked: "I don't think we need to validate the fact that the two rays intersect here. This is…focusing on *minutia* that will prevent kids from focusing on the important parts of the problem (participant ID 2333, emphasis added)." Yet other open responses indicate that the scrutiny of some aspects of a proof is compulsory. For example, a participant who viewed story B—one in which the teacher allows a student to make statements about the sum of the angles of a triangle as opposed to the sum *of the measures* of the angles—said: "The teacher is down-playing the little things. Sometimes those little things can change the whole outcome (participant ID 2300)." Viewing this same story, a different participant commented: "When you do proofs, you *can't assume anything* (participant ID 2359, emphasis added)." These comments would seem to be directly at odds with those reported above. That both under-scrutiny of the written argument (second example responses) of a proof and hyper-scrutiny of the diagram accompanying a proof (first example responses)—practices that could be seen as equivalent from the perspective of justifying every step in a proof—can draw the concern of secondary teachers provide evidence that the routines for checking the details of a prof are, in fact, norms.

Two-column proof has been criticized for being ritualistic or attentive to excessive detail (e.g., Harel & Sowder, 1998; Schoenfeld, 1988); however, our research shows that such statements are too broad—attention to detail depends on what details are being considered and how those details are being expressed. When it comes to statements—such as the existence of a point of intersection—that are tacitly warranted by a diagram, participants reacted unfavorably to episodes that showed a teacher asking for the explicit justification that would warrant those statements, on the grounds that doing so was focusing on minutia. However, when it comes to statements—such as deducing the congruence of two segments from the definition of midpoint—that are tacitly entailed by definitions, participants reacted unfavorably to episodes that showed a teacher *not* asking for the explicit justification that would warrant those statements, on the grounds that every step in a proof requires an explicit justification. That teachers would hold different views of the appropriate level of detail in a proof is not *a priori* obvious, and the account we have provided highlights the affordances of the method we have developed.

CONCLUSION

The research reported here describes a method for confirming that a routine classroom practice is a norm and uses that method to confirm the existence of semiotic norms when doing proofs in geometry. The articulation of a semiotic norm contributes an elaboration of the theory of instructional exchanges, while its experimental confirmation contributes a method that can be used to identify normative practices in instruction. More generally, we have shown that representations of lessons may be used in an experimentally controlled way to target what teachers notice about instruction.

References

Christie, F. (1997). Curriculum macrogenres as forms of initiation into a culture. In F. Christie & J. R. Martin (Eds.), *Genre and institutions: Social processes in the workplace and school* (pp. 134-160). London: Cassell.

Garfinkel, H. (1963). A conception of and experiments with “trust" as a condition of concerted stable actions. In J. O’Brien (Ed.), *The production of reality: Essays and readings on social interaction* (pp. 381-392). London: Sage.

Harel, G., & Sowder, L. (1998). Students’ proof schemes: Results from exploratory studies. *Research in Collegiate Mathematics Education III*, *7*, 234-282.

Herbst, P. G. (2006). Teaching geometry with problems: Negotiating instructional situations and mathematical tasks. *Journal for Research in Mathematics Education, 37*, 313-347.

Herbst, P., & Chazan, D. (2003). Exploring the practical rationality of mathematics teaching through conversations about videotaped episodes: The case of engaging students in proving. *For the Learning of Mathematics*, *23*(1), 2-14.

Herbst, P., & Chazan, D. (2011). Research on practical rationality: Studying the justification of actions in mathematics teaching. *The Mathematics Enthusiast*, *8*(3), 405-462.

Herbst, P., & Chazan, D. (2012). On the instructional triangle and sources of justification for actions in mathematics teaching. *ZDM*, *44*(5), 601-612.

Herbst, P., & Miyakawa, T. (2008). When, how, and why prove theorems? A methodology for studying the perspective of geometry teachers. *ZDM*, *40*(3), 469-486.

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex Publishing Corporation.

Martin, J. R., & Rose, D. (2008). *Genre relations: Mapping culture*. London: Equinox.

Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. Cambridge, MA: Harvard University Press.

Nachlieli, T., Herbst, P., & González, G.(2009). Seeing a colleague encourage a student to make an assumption while proving: What teachers put to play in casting an episode of geometry instruction. *Journal for Research in Mathematics Education*, *40*(4), 427-459.

Santagata, R., & Stigler, J. W. (2000). Teaching mathematics: Italian lessons from a cross-cultural perspective. *Mathematical Thinking and Learning*, *2*(3), 191-208.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of 'well-taught' mathematics courses. *Educational Psychologist*, *23*(2), 145-166.

Stigler, J. W., & Hiebert, J. (2009). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Simon & Schuster.

Van Leeuwen, T. (2004). *Introducing social semiotics: An introductory textbook*. New York: Routledge.